Photometric Redshift estimation using Gaussian Process Regression

Michael Way (NASA/Goddard Institute for Space Studies)
Paul Gazis, Jeffrey Scargle (NASA/Ames, Space Sciences Division)
Ashok Srivastava (NASA/Ames Intelligent Systems Division)
Les Foster + Students (San Jose State University)

http://astrophysics.arc.nasa.gov/~mway/ETH-201107.pdf

ETH 2011/07

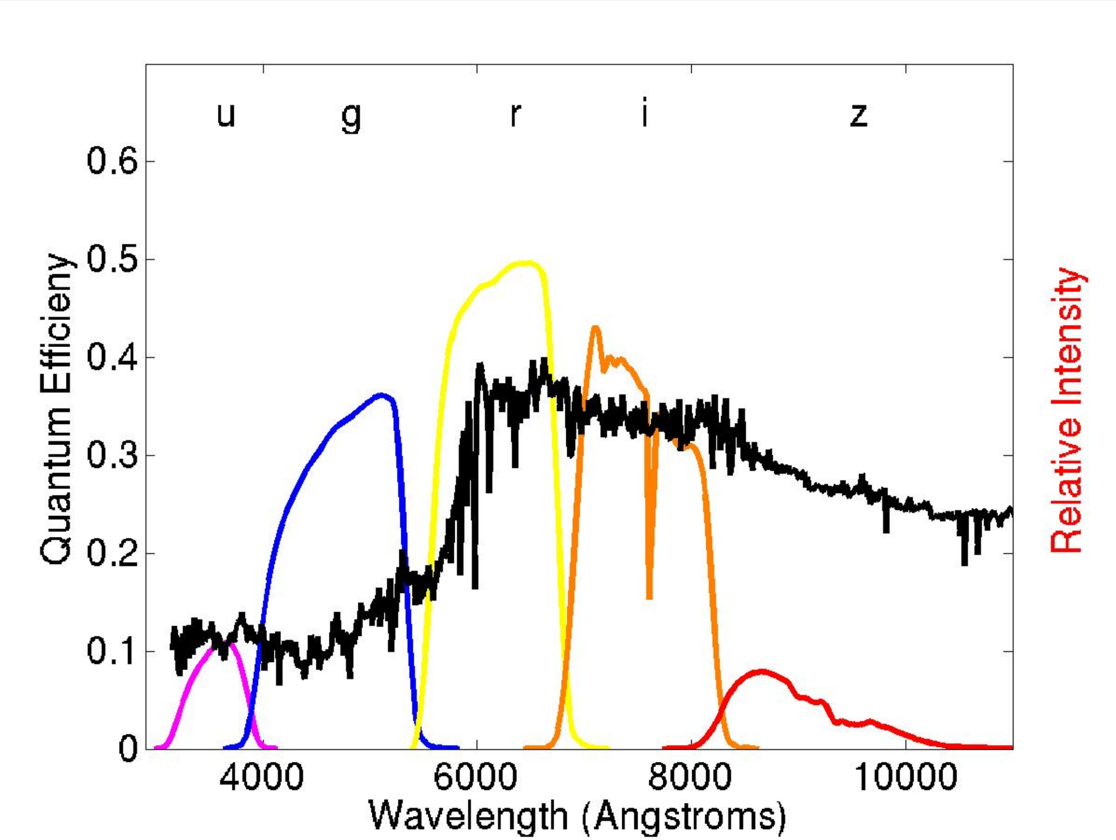
Outline

- What are Photometric Redshifts?
- Common training set methods
- What is Gaussian Process Regression?
- Do different kinds of Kernels matter?
- Matrix Inversion Options
- How many galaxies do I need to get a good fit?
- Do SDSS morphological indicators help?
- Do SDSS + 2MASS colors really help?

What are Photometric Redshifts?

Photometric Redshifts: A **rough** estimate of the redshift of a galaxy without having to measure a spectrum.

$$Z_{spec} = (\lambda_{measured} - \lambda_{rest})/\lambda_{rest} \qquad z_{photo} = z(C, m)$$

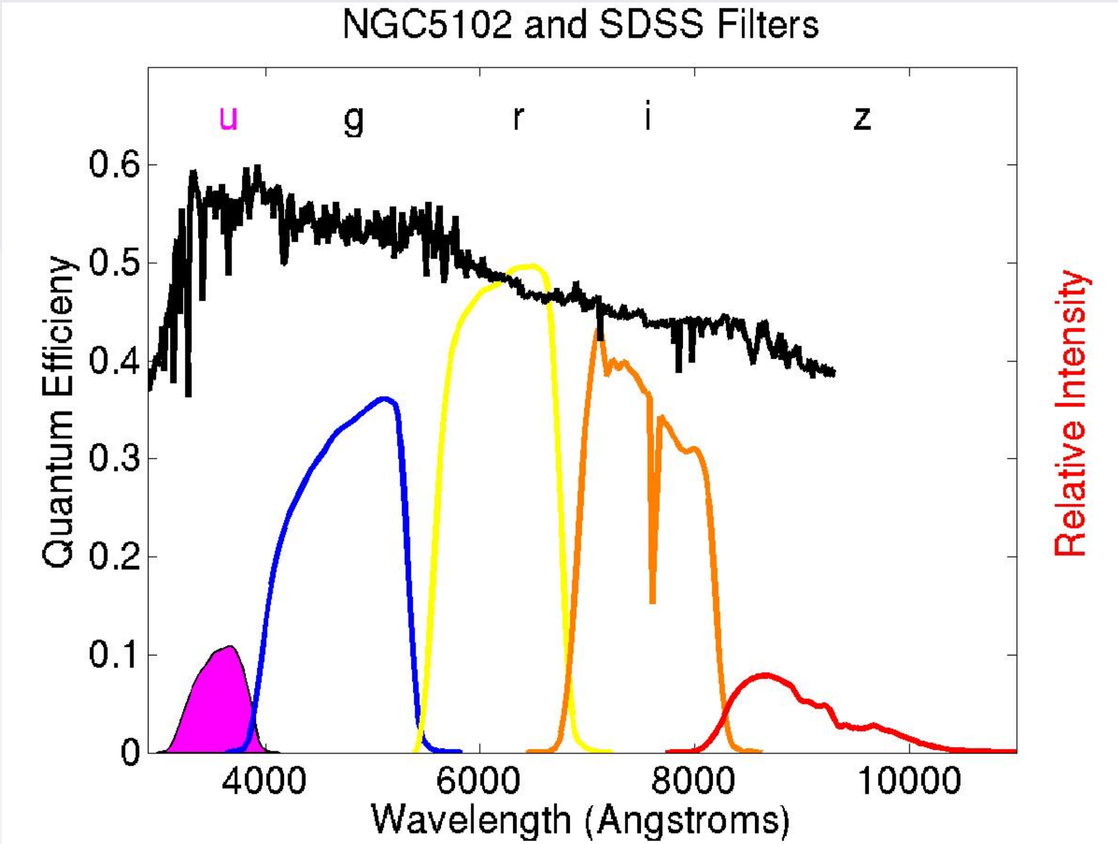


What are Photometric Redshifts?

$$Z_{spec}=(\lambda_{measured}-\lambda_{rest})/\lambda_{rest}$$

$$z_{photo}=z(C,m)$$

z=0.0

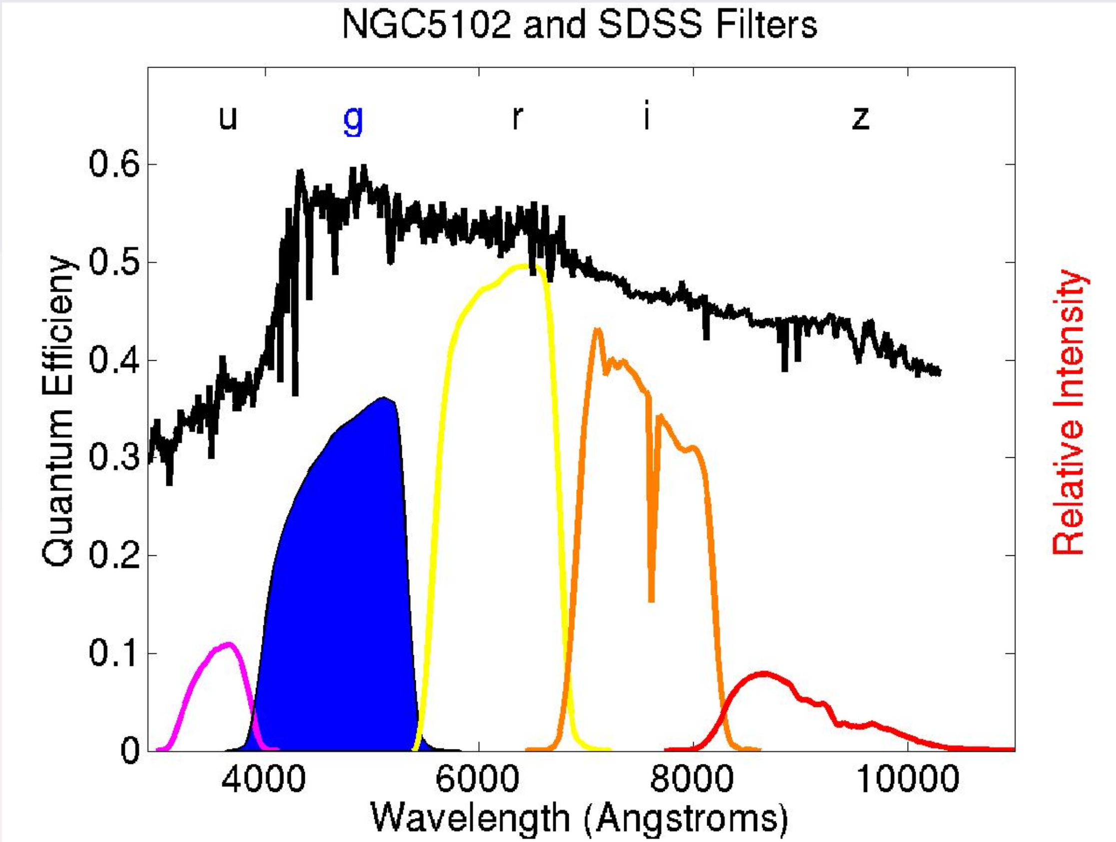


What are Photometric Redshifts?

$$Z_{spec}=(\lambda_{measured}-\lambda_{rest})/\lambda_{rest}$$

$$z_{photo}=z(C,m)$$

z~0.06 (18000 km/s)

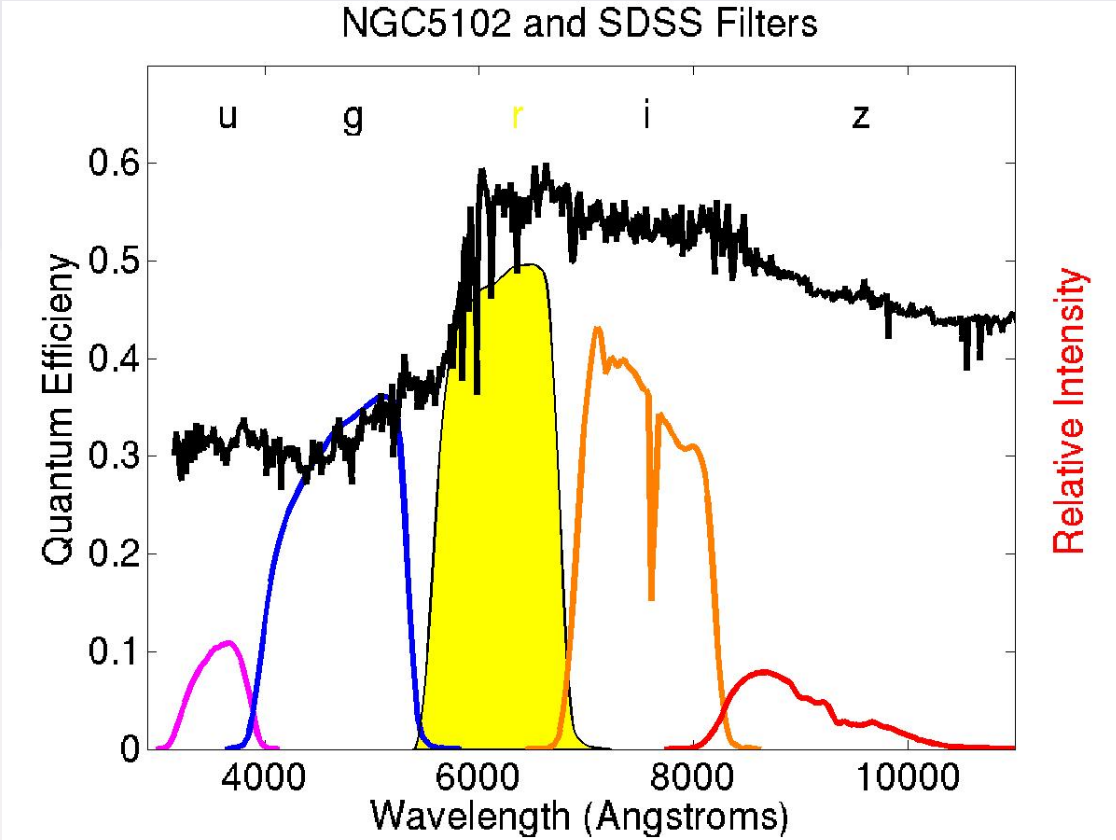


What are Photometric Redshifts?

$Z_{spec} = (\lambda_{measured} - \lambda_{rest}) / \lambda_{rest}$

$z_{photo} = z(C,m)$

z~0.6



What are Photometric Redshifts?

$\mathbf{Z_{spec}=(\lambda_{measured}-\lambda_{rest})/\lambda_{rest}}$ $z_{photo}=z(C,m)$

z~0.90

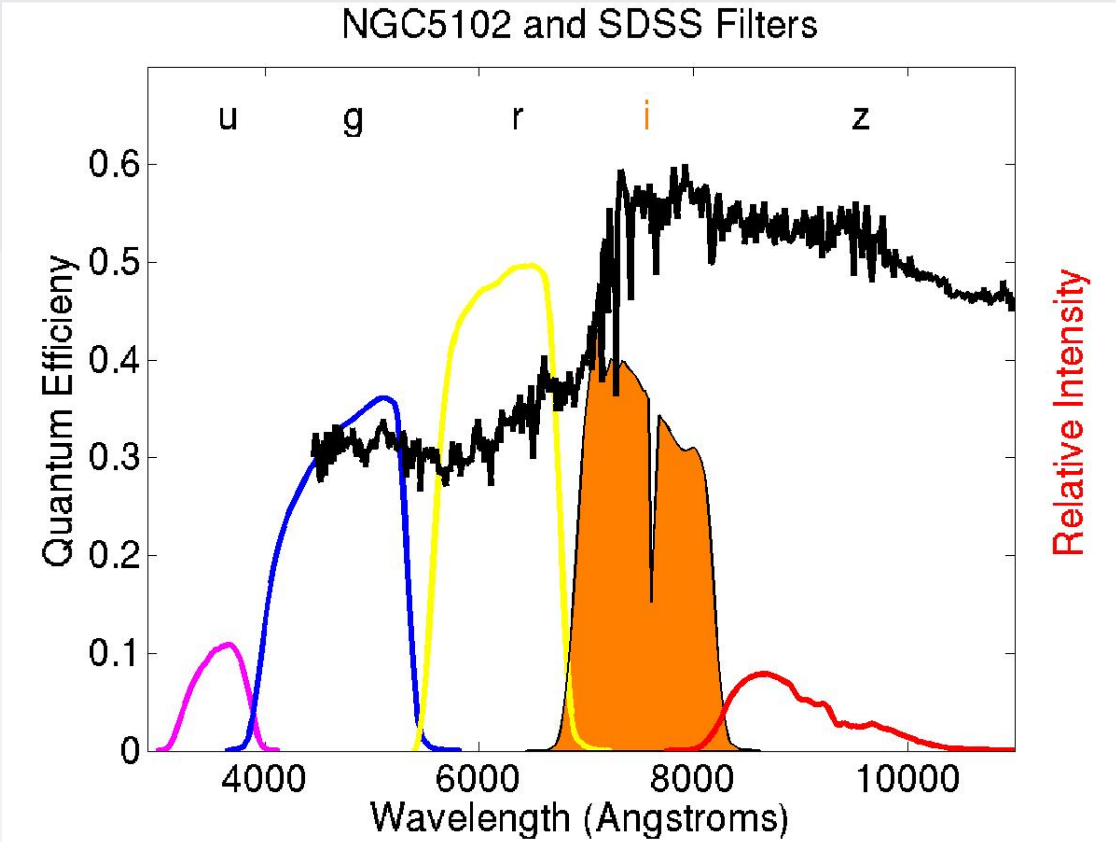


Photo-z methods

1.) Spectral Energy Distribution (SED) Fitting:

- model based approach
- uses redshifts derived from spectra of artificial galaxies (e.g. Bruzual & Charlot)

2.) Training-Set methods:

- **empirical approach**
- **uses *spectroscopic* redshifts from a sub-sample of galaxies with the same band-pass filters**

Photo-z The Empirical Approach

Training Set Methods need a sub-sample of Galaxies:

- of known spectroscopic redshift
- with a comparable range of **magnitudes** (u g r i z) to our Photometric survey objects
  For the SDSS MGS that is r<17.77 (NOT 17.77<r<22)
- These will be our "Training Samples"

"Training Set" Methods

Galaxy Photometric Redshift Prediction History

u-g-r-i-z <–> redshift

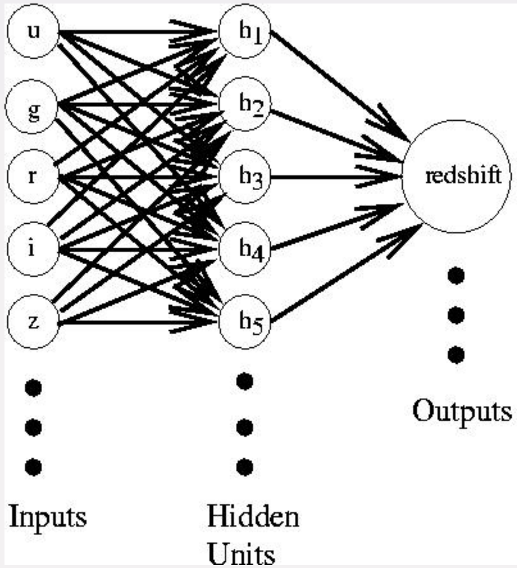
- Linear Regression was first tried in the 1960s
- Quadratic & Cubic Regression (1970s)
- Polynomial Regression (1980s)
- Neural Networks (1990s)
- Kd Trees & Bayesian Classification Approaches (1990s)
- Support Vector Machines & GP Regression (2000s)

Gaussian Process Regression fitting

Gaussian Process Regression ⇔ Kernel Methods

Kernel Methods have replaced Neural Networks in the Machine Learning literature

WHY?: given a large # of hidden units => GP (Neal 1996).



$h_n > 100$

→→→→→

Kernel Methods - Gaussian Process Regression

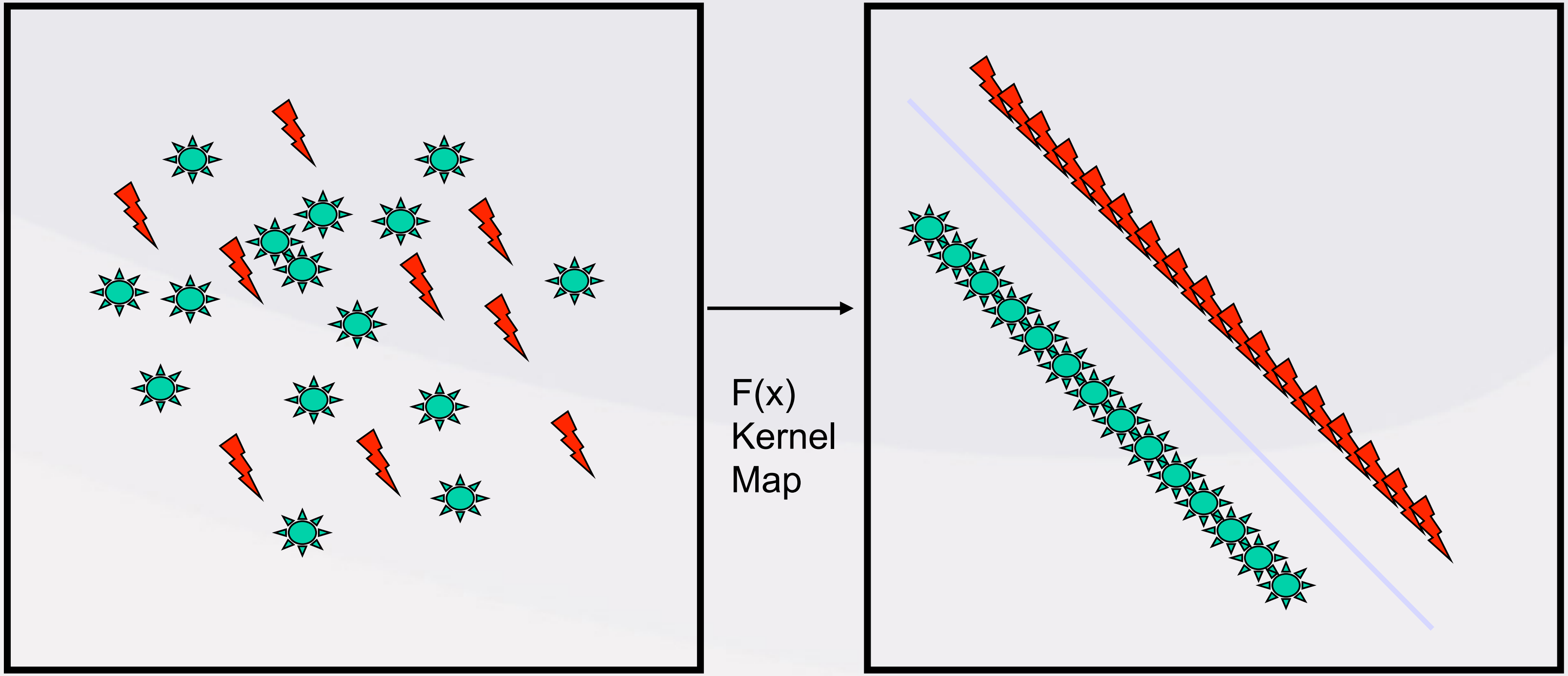
GP regression builds a linear model in a very high dimensional *parameter space* ("feature space" → Hilbert space).

- One can map the data using a function F(x) [kernel] into this high (or infinite) dimensional *parameter space* where one can perform linear operations.

The value of kernels

Original Data without Kernel

Mapped Data using Kernel



Data in original space: highly complex decision boundaries.

Data in high dimensional feature space after mapping through F(x) can yield simple decision boundaries.

GP Regression: Advantages

GP Advantages:

- Small input data training samples yet low errors
- Realistic estimation of individual redshift errors

GP Regression: Problems?

GP Disadvantages:

1.) Possibly large CPU time requirements

– The Kernel (Covariance Matrix) **can** be large: $K=(\lambda^2 I+XX^T)^2$ if $X=5\text{x}180{,}000$ (our case) then K is a matrix 180,000 x 180,000 and we have:

$$y^* = K^* (\lambda^2 I + K)^{-1} y$$

– Need to invert this large (non-sparse) K matrix

- $O(N^3)$ operation, $O(N^2)$ memory

2.) Kernel Selection is ambiguous?

GP: Which Kernel??

Kernel Selection: Pick a transfer/covariance function

Matern Class Fcn

$$k(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu r}}{l}\right)^{\nu} J_{\nu}\left(\frac{\sqrt{2\nu r}}{l}\right)$$

$$\nu \rightarrow \infty$$

Radial Basis Fcn

$$k(r) = \exp\left(\frac{r^2}{2l^2}\right)$$

Rational Quadratic

$$k_{RQ}(r) = 1 + \left(\frac{r^2}{2\alpha l^2}\right)^{-\alpha}$$

Polynomial

$$k(x,x') = \left(\sigma_o^2 + x^T \sum_p x'\right)^p$$

Neural Nets

$$k_{NN}(x,x') = \frac{2}{\pi}\sin^{-1}\left(\frac{2x^T\Sigma x'}{\sqrt{(1+2x^T\Sigma x)(1+2x'^T\Sigma x')}}\right)$$

3 options for matrix Inversion

Option 1: Take a random sample of ~1000 galaxies & invert that while bootstrapping n times from full sample (Paper I)

- Advantages
  - Can run on a 32bit computer
  - Doesn't take too long: $O(N^3)$ operation
  - Doesn't take up too much memory $O(N^2)$
- Disadvantages
  - Accuracy suffers – we don't sample enough galaxies/SEDs

GP Matrix Inversion

Option 2: Use a 64 bit SSI computer

- Advantages
  - Accuracy – we invert the full matrix using all sample galaxies
- Disadvantages
  - Takes a VERY long time: $O(N^3)$ operation
  - We need a lot of memory: $O(N^2)$
  - Hard to get access to such a computer for such a long time
    - e.g. Mac Pro: 64 bit, 4 cpu, 16GB of RAM, max is ~20000x20000 in Matlab

GP Matrix Inversion

Option 3: Low-rank matrix approximations: Subset of Regressors, Cholesky Decomposition, Projected Process Approximation, etc.

(Paper II: https://dashlink.arc.nasa.gov/algorithm/stablegp)

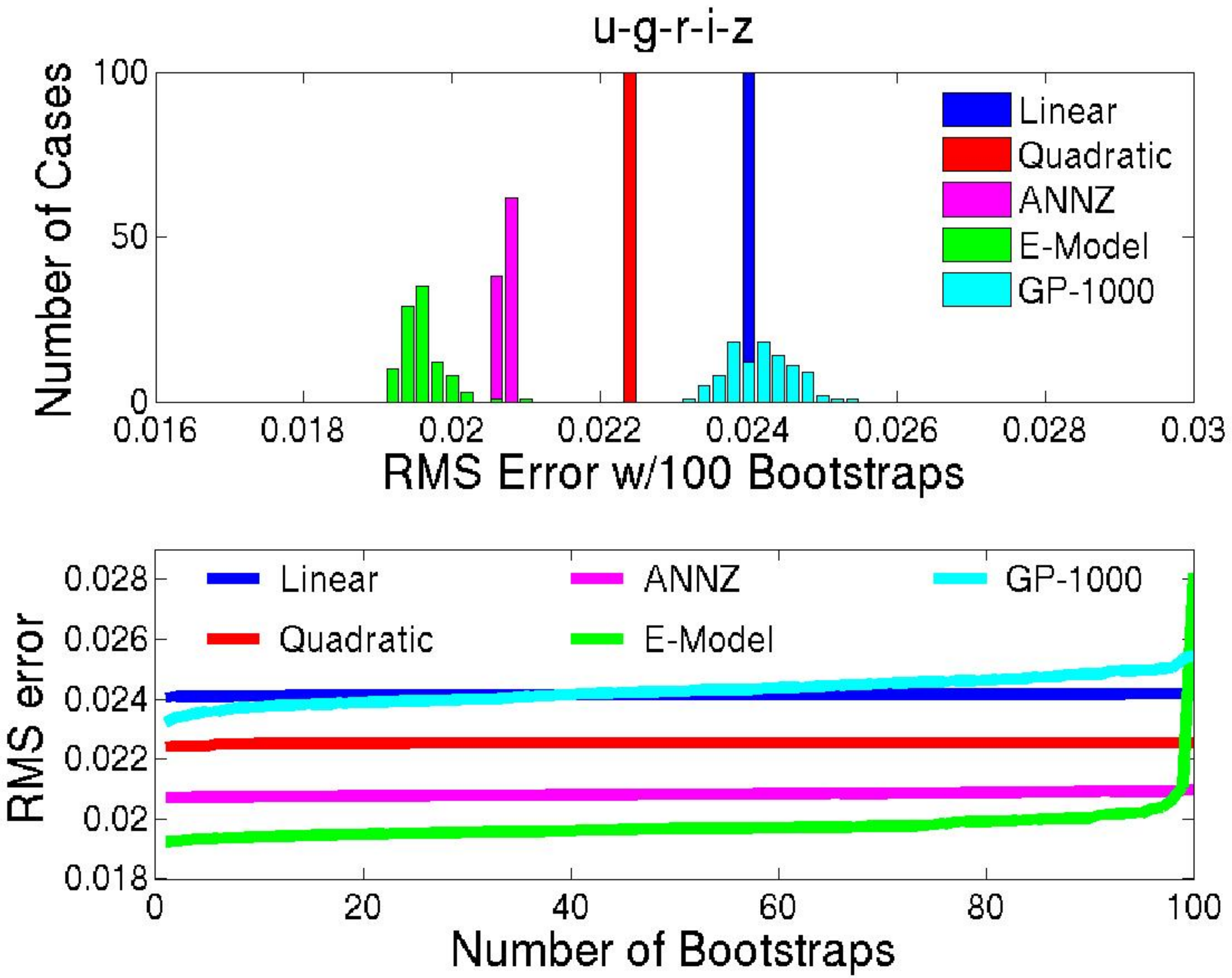
- Advantages
  - Accuracy – we invert much more of the full matrix
  - Doesn‘t take too long: dependent upon rank=N
  - Doesn‘t take up too much memory
- Disadvantages
  - Hard to know how it compares to full matrix inversion

GP Regression (Results)

Results: SDSS (DR3) Main Galaxy Sample

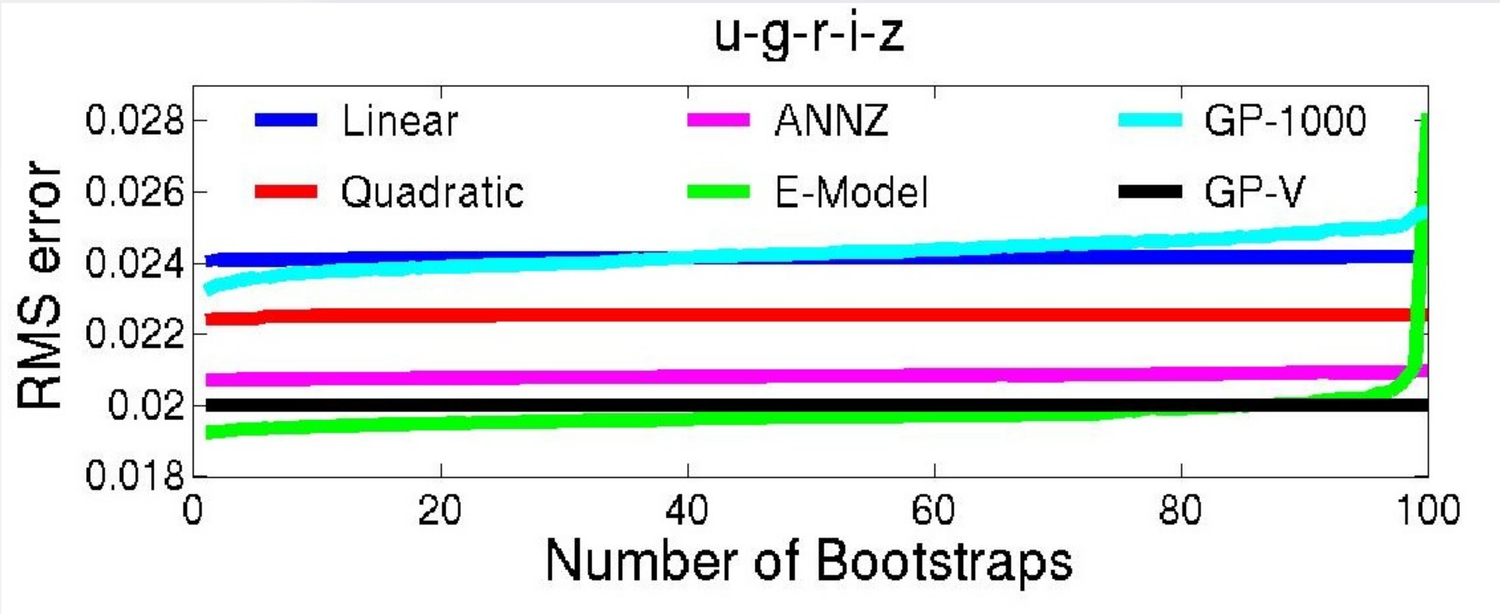
- Paper I: Compared linear, quadratic, Neural Networks and GPs on the SDSS-DR3
  - With ONLY 1000 samples GPs performed well compared to the other methods
- Paper II: Low-rank matrix inversion approximations with more appropriate Kernel
  - GPs performed better than all other methods to date

Paper I Results: Comparing Methods



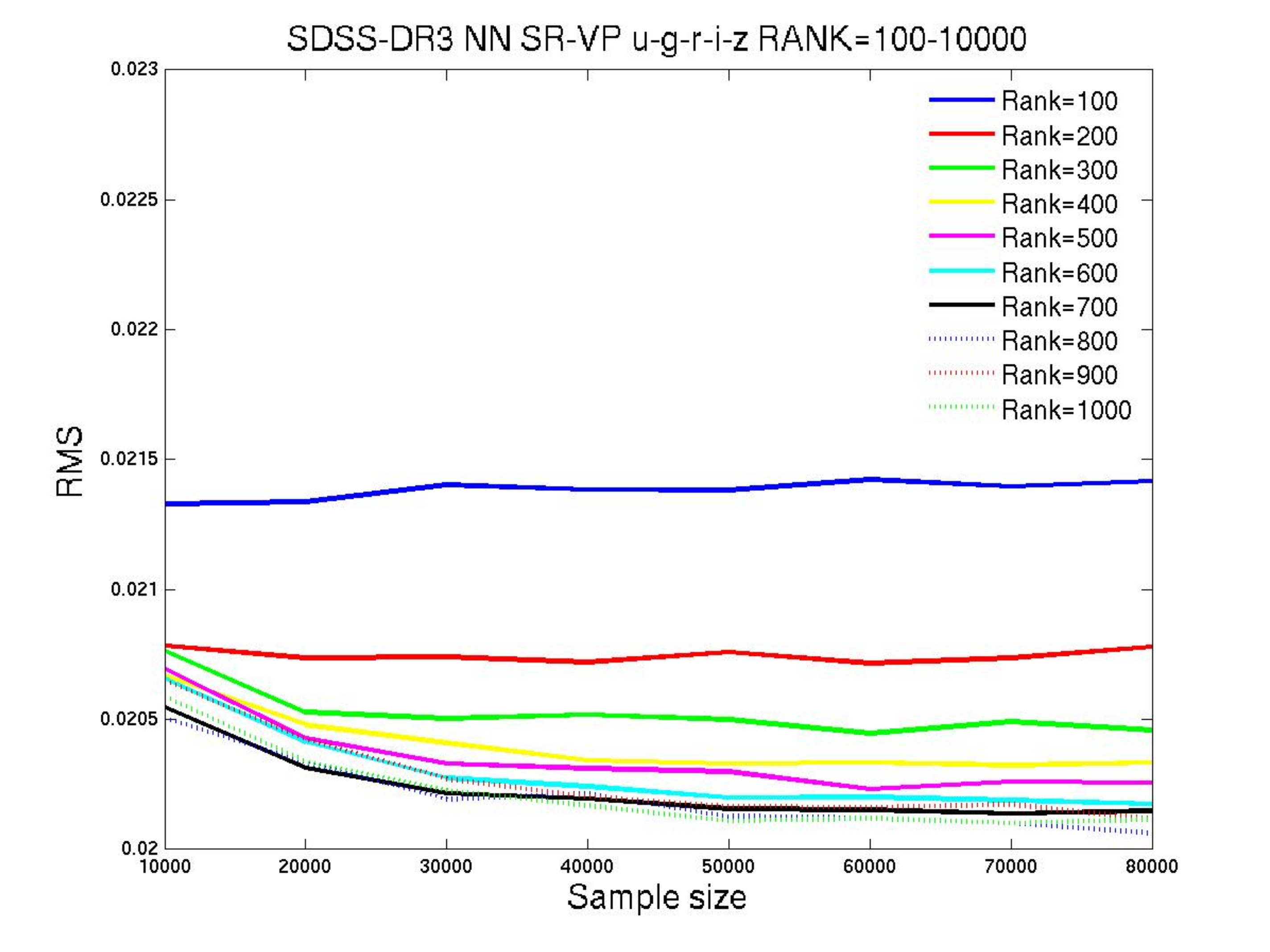
New Results: Paper II

- GPR with rank=1000 : V-method : Polynomial Kernel
- Better results possible using VP method & NN Kernels



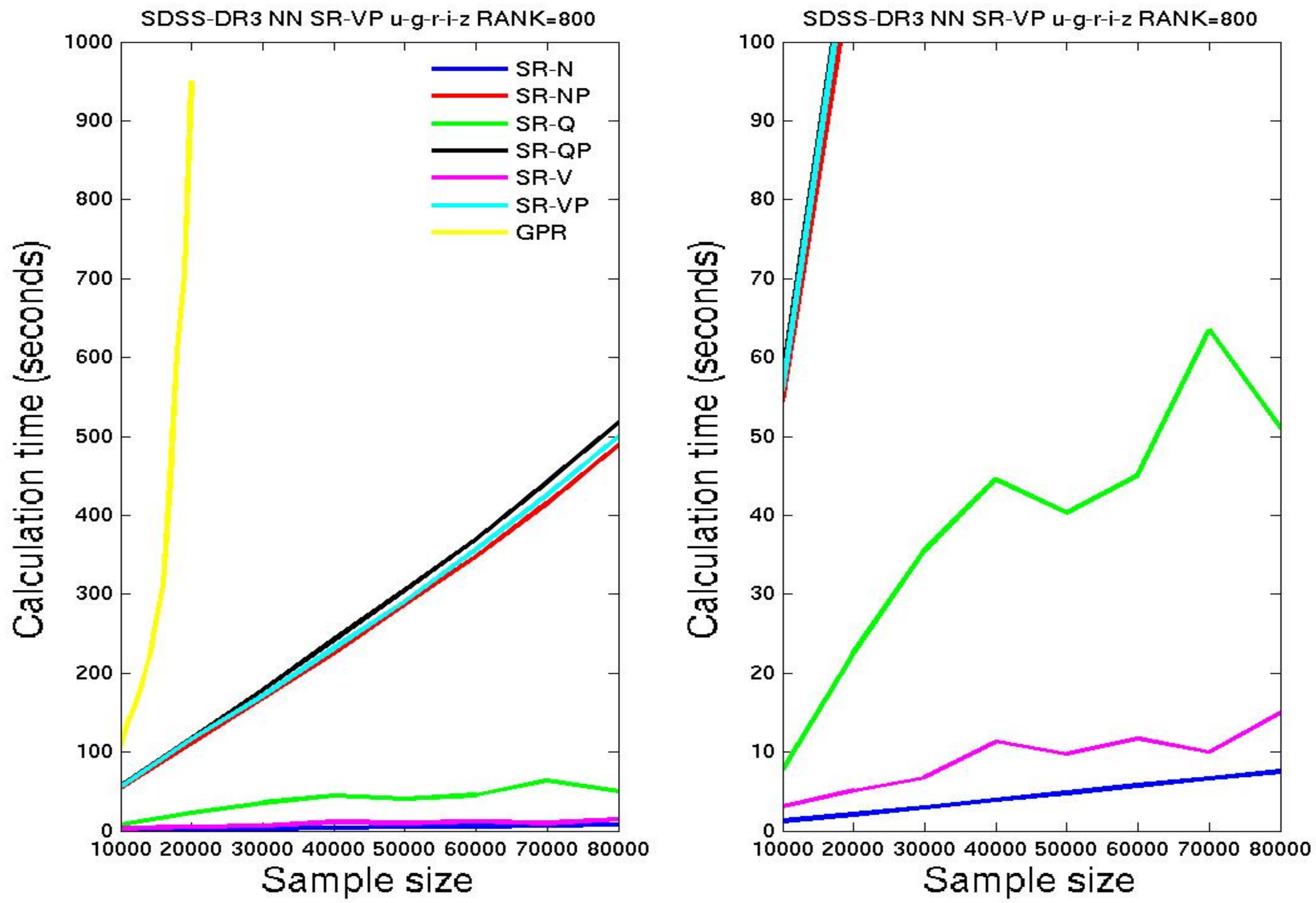
Rank vs Sample Size

Near optimal is ~40,000 samples, Rank=800



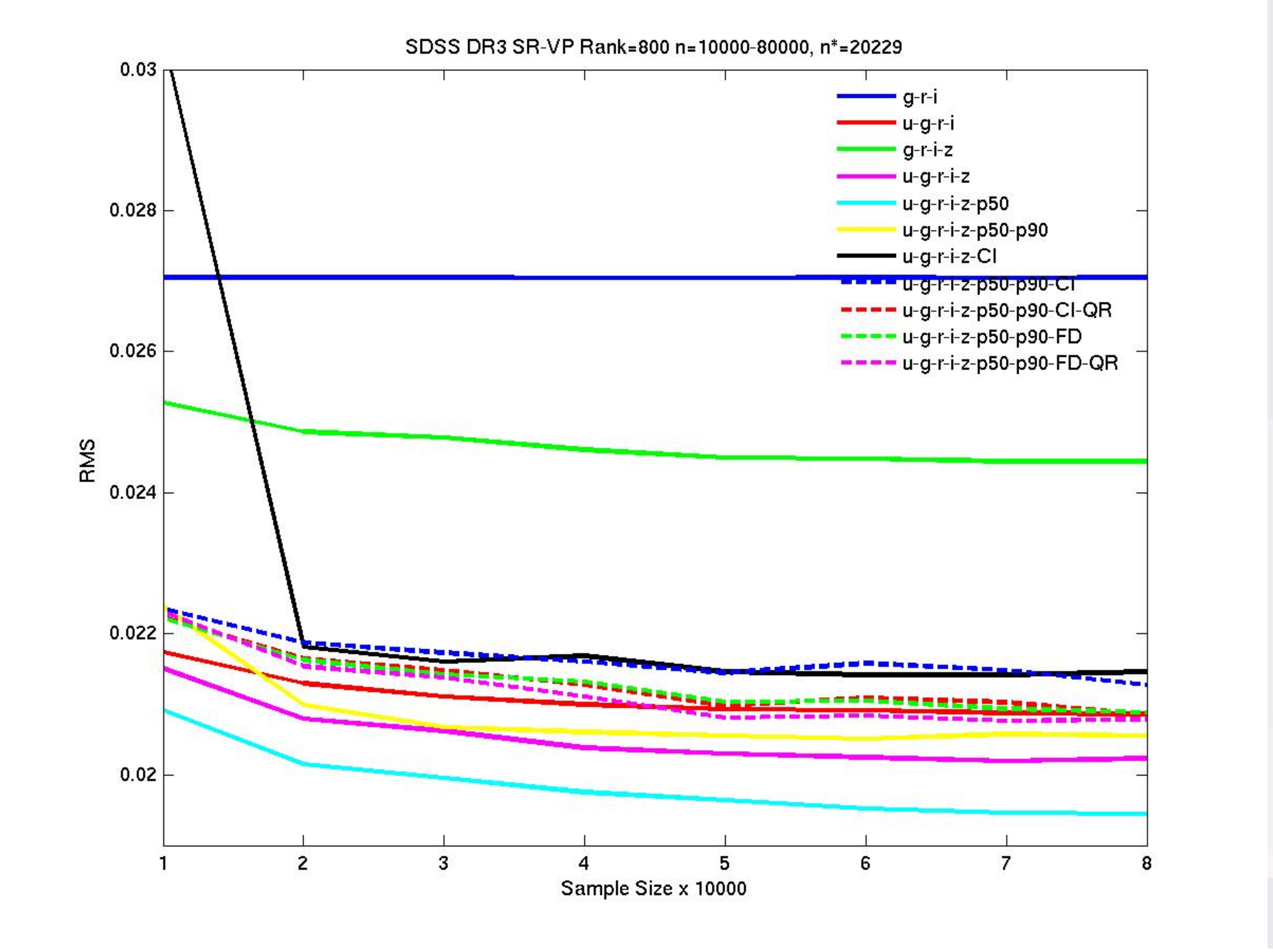
Calculation Time?!

Matrix Inversion: that $O(N^3)$ business?



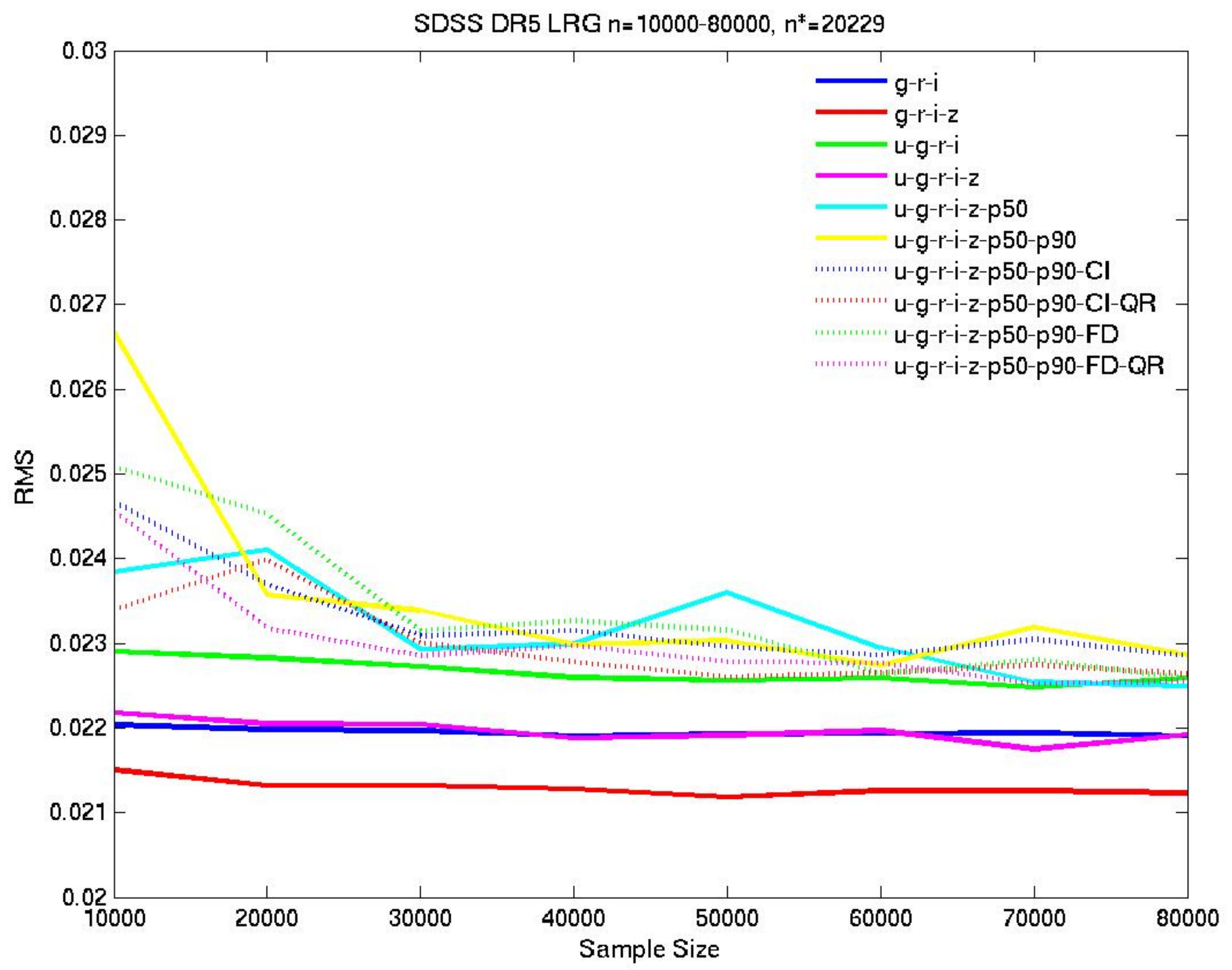
Secondary isophotal parameters?

SDSS-DR3 Main Galaxy Sample



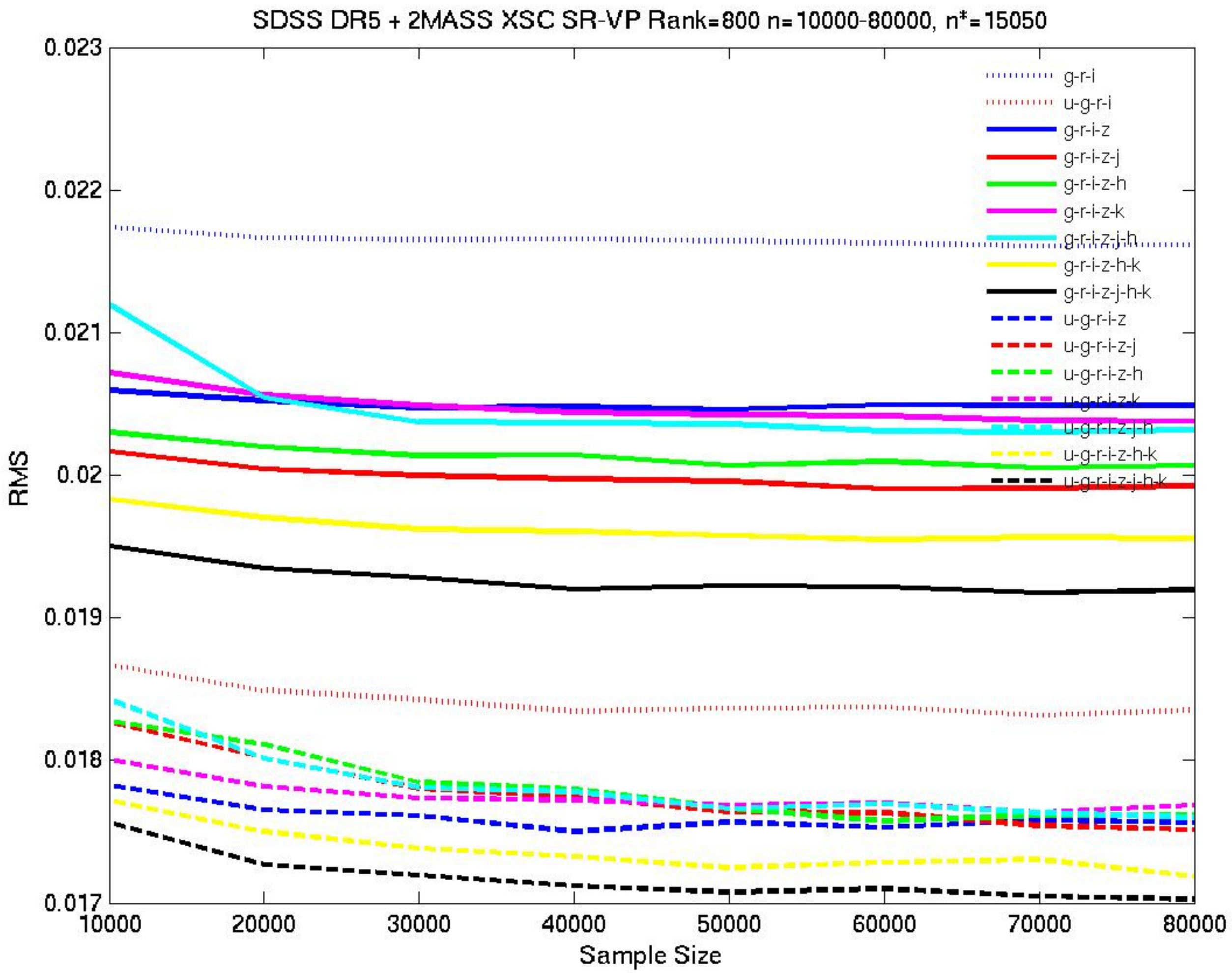
Secondary isophotal parameters?

SDSS-DR5 Luminous Red Galaxy Sample



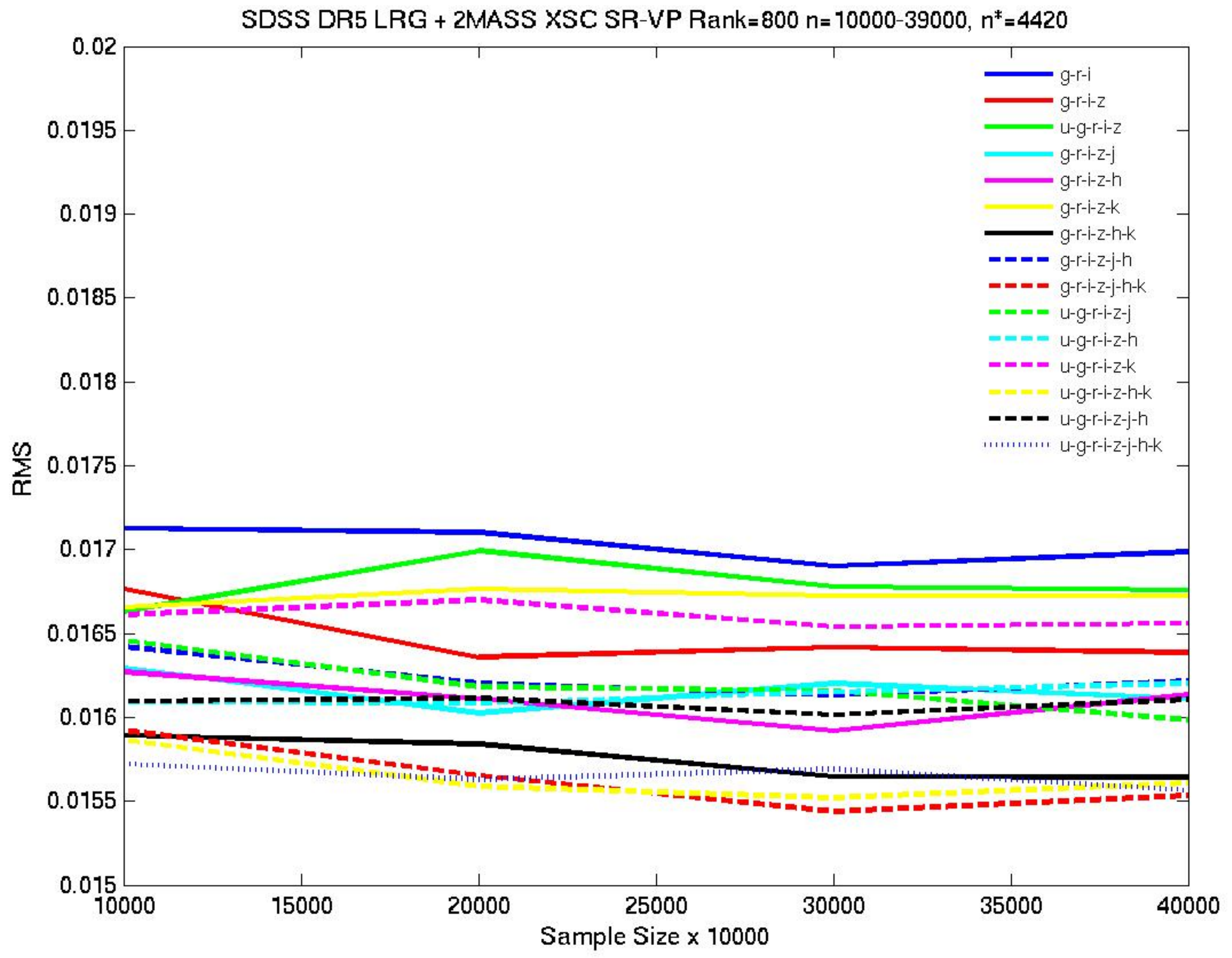
SDSS-MGS + 2MASS xsc

SDSS-DR5 MGS + 2MASS



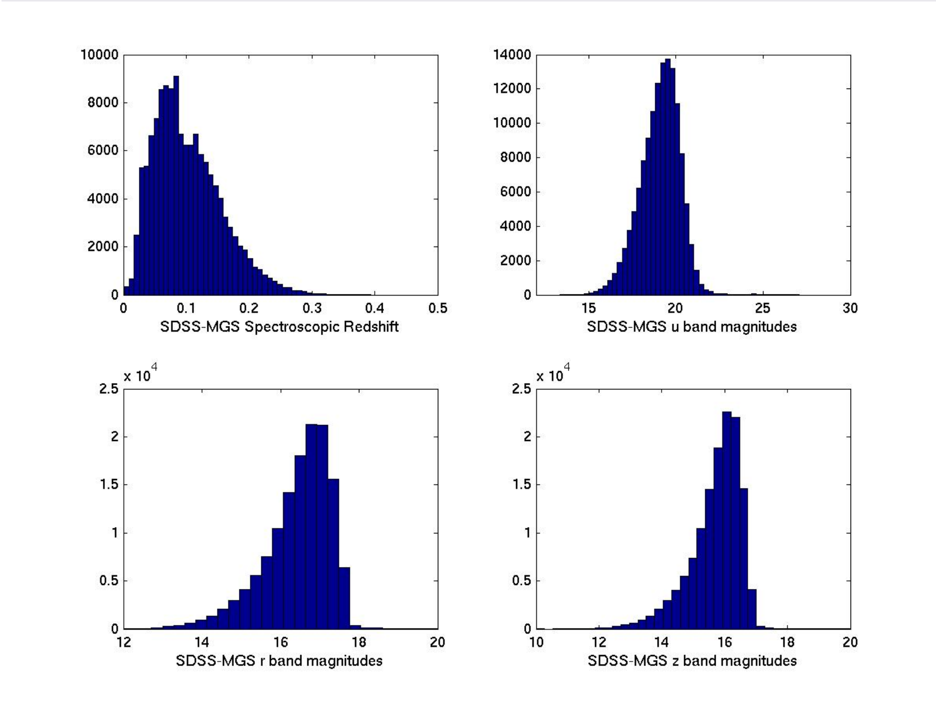
SDSS-LRG + 2MASS xsc

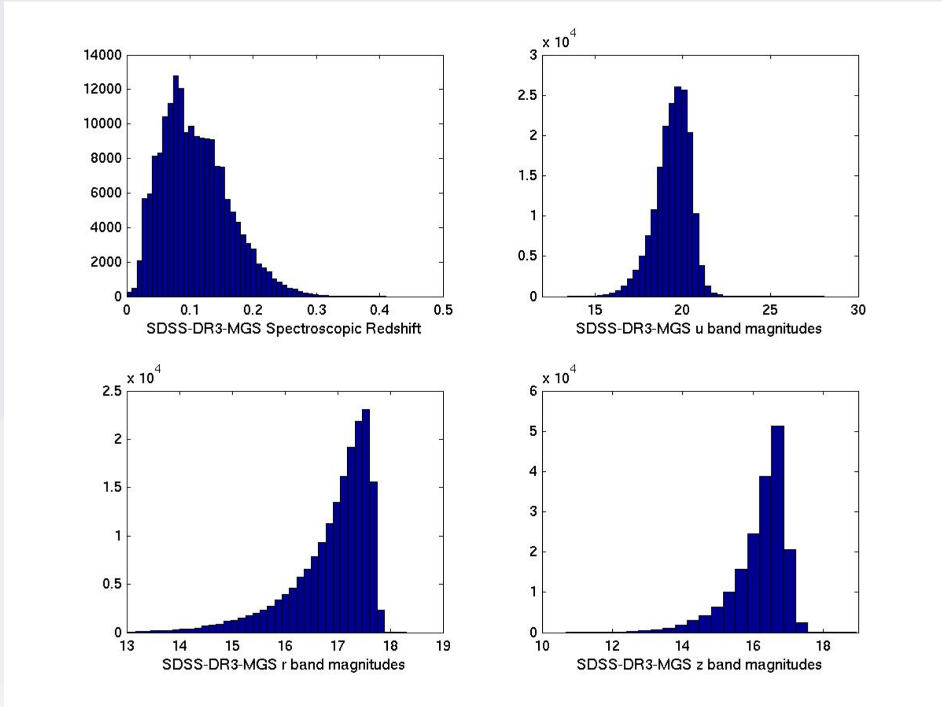
SDSS-DR5 LRG + 2MASS



SDSS-MGS + 2MASS xsc

u-g-r-i-z magnitudes are suddenly better?



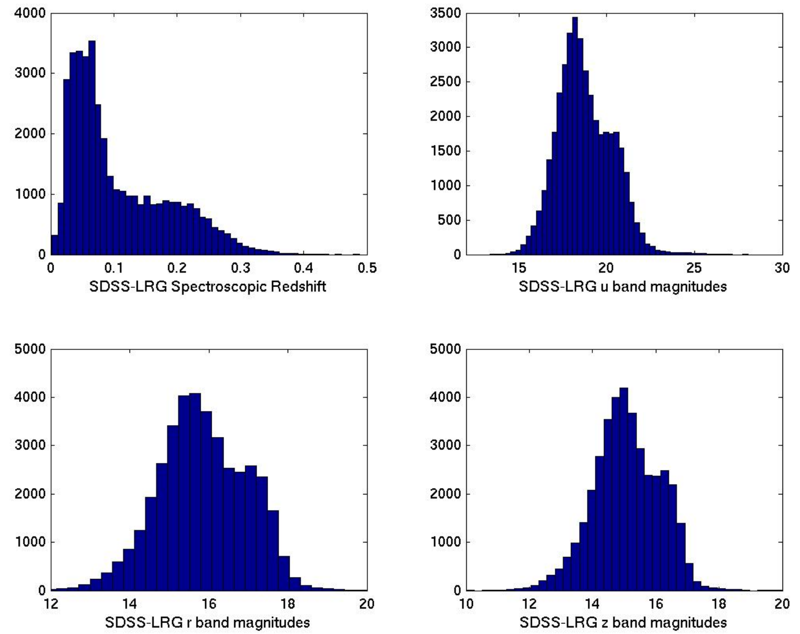


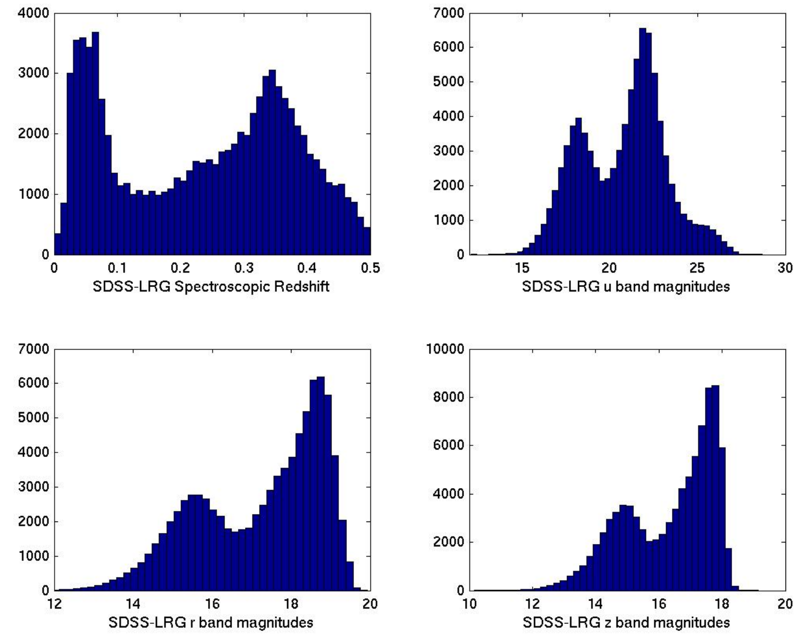
2MASS/SDSS-DR5 match

SDSS-DR5 only

SDSS-LRG + 2MASS xsc

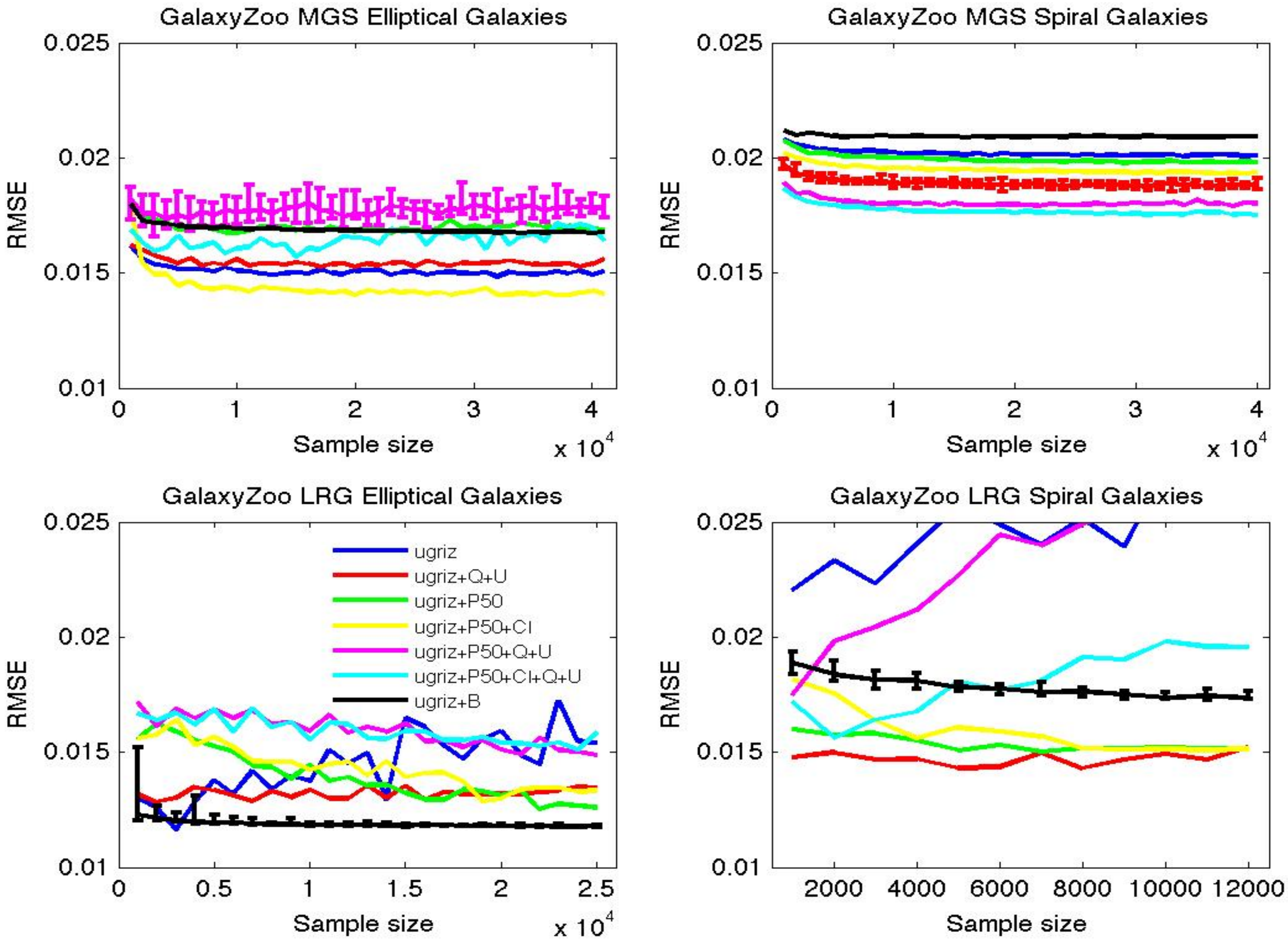
u-g-r-i-z magnitudes are suddenly better?





2MASS+SDSS/LRG-DR5 SDSS/LRG-DR5 only

SDSS + GZ Morphology?



Results?

- GPR is now faster & more competitive
- ~40,000 objects are required for optimal results when using the SDSS-MGS, while LRG sample is good at 10,000
- Additional Near IR filters (2MASS) help?
- Secondary isophotals work: MGS vs LRG
- GalaxyZoo morphology makes a difference

Thanks

This was made possible by the cooperation of Earth Scientists, Astronomers, Machine Learning people and Mathematicians in 4 different groups.

Thanks to:

- Les Foster & students (San Jose State University)
- Ashok Srivastava (NASA/Ames, Intelligent Systems)
- Rama Nemani (NASA/Ames, Earth Science)
- Paul Gazis & Tim Lee (NASA/Ames, Space Science)